BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Exhibit 1

J. gen. Virol. (1987), **68**, 615–626. *Printed in Great Britain*

Key words: *BNYVV/genome organization/sequence homology*

Nucleotide Sequence of Beet Necrotic Yellow Vein Virus RNA-1

By S. BOUZOUBAA, L. QUILLET, H. GUILLEY, G. JONARD AND K. RICHARDS*

Laboratoire de Virologie, Institut de Biologie Moléculaire et Cellulaire, 15 rue Descartes, 67084 Strasbourg Cédex, France

(*Accepted 4 November 1986*)

SUMMARY

The complete nucleotide sequence of beet necrotic yellow vein virus RNA-1 is presented. The RNA molecule is 6746 nucleotides long excluding the poly(A) tail and has one long open reading frame encoding a polypeptide of M_r 237 389. The 3′ terminal 60 residues of BNYVV RNA-1 display extensive sequence homology with the corresponding portions of BNYVV RNA-2, -3 and -4. Additional 3′ terminal homology exists between RNA-1 and -2. The sequence of the M_r 237389 RNA-1-encoded polypeptide shares domains of amino acid homology with polypeptides thought to be involved in replication of RNA from tobacco mosaic virus and several other viruses. Amino acid sequence homologies between two open reading frames of BNYVV RNA-2 and two frames of RNA-β from barley stripe mosaic virus have also been detected.

INTRODUCTION

Beet necrotic yellow vein virus (BNYVV) is a multicomponent soil-borne rod-shaped virus responsible for a severe disease of sugarbeet called rhizomania (Tamada, 1975). Field isolates of BNYVV typically contain four single-stranded 5′-capped and 3′-polyadenylated plus-sense RNA molecules (Putz, 1977; Putz *et al.*, 1983). BNYVV RNA-2 [4612 nucleotides, excluding the poly(A) tail], RNA-3 (1774 nucleotides) and RNA-4 (1467 nucleotides) have already been sequenced (Bouzoubaa *et al.*, 1985, 1986). In this paper, we report the complete sequence of 6746 nucleotides for RNA-1, thus completing the molecular description of the BNYVV genome. Similarities of sequence and genetic organization between BNYVV and other RNA viruses are described.

METHODS

cDNA synthesis and cloning. RNA-1-specific cDNA clone pBF5 was prepared from RNA of BNYVV isolate F2 as described previously (Richards *et al.*, 1985). Other RNA-1-specific cDNA clones were prepared with RNA from isolate F13, a subisolate of F2 (Ziegler *et al.*, 1985). In this case the primers for first-strand cDNA synthesis were synthetic oligodeoxyribonucleotide primers 15 to 22 residues in length complementary to portions of RNA-1 close to the 5′ limit of the known sequence. The oligodeoxyribonucleotides were synthesized with an Applied Biosystems 381A apparatus and purified by polyacrylamide gel electrophoresis. Clones pBF51 and pBF5201 were obtained by directly cloning dC-tailed RNA–cDNA hybrids into *Pst*I-linearized dG-tailed pUC9 (Bouzoubaa *et al.*, 1986). Clones pBF5220, pBF541 and pBF542 were obtained by treating the hybrid with RNase H (Bethesda Research Laboratories) and DNA polymerase I (Boehringer) (Okayama & Berg, 1982; Gubler & Hoffman, 1983) prior to dC-tailing and cloning. For clone pBF553 second-strand synthesis was primed by a synthetic oligodeoxyribonucleotide corresponding to nucleotides 1 to 23 of RNA-1.

Recombinant clones were screened for large inserts by restriction enzyme analysis and the selected clones were screened for RNA-1-specific sequences by Northern hybridization to viral RNA using nick-translated plasmid DNA from candidate clones as probe (Bouzoubaa *et al.*, 1985). This test proved necessary to eliminate clones containing RNA-2 and -3 specific cDNA which occurred frequently in spite of the fact that care was taken to choose primer sequences with a minimum of complementarity with the known RNA-2 and -3 sequences.

Sequence analysis. cDNA inserts were sequenced by partial chemical degradation (Maxam & Gilbert, 1977) after digestion with appropriate restriction enzymes and ^{32}P end-labelling or by subcloning restriction fragments into M13mp10 or M13mp11 (Messing, 1983) for sequencing by the dideoxyribonucleotide chain termination

method (Sanger *et al.*, 1977). Sequence results were also obtained by primer extension of 5′-^{32}P-labelled synthetic primers hybridized to RNA-1 and analysis of the 5′-labelled cDNA transcript by partial chemical degradation (Bouzoubaa *et al.*, 1986). The sequence was determined on both strands over about 75% of its total length.

Data were manipulated with UWGCG programs (Devereux *et al.*, 1984) on a VAX 11/750 minicomputer. DNA and amino acid sequence homologies were compared by the matrix method (Maizel & Lenk, 1981) using the programs COMPARE and DOTPLOT. Amino acid homology scoring was with the modified MDM78 matrix (Staden, 1982) normalized by multiplication by 0·1. Stringencies of 32·0 to 33·0 were used for a 30 amino acid window. Homologous regions were aligned with the program BESTFIT using a direct identity matrix for DNA comparisons (gap penalty 5, gap length penalty 0·3) and the log odds matrix for 250 accepted point mutations (normalized by multiplication by 0·25) of Dayhoff *et al.* (1978) for polypeptides. In the latter case a gap penalty of 4 and gap length penalty of 0·15 were used. Quality was calculated as number of matches − (0·05 × number mismatches) − (gap penalty × number of gaps) − (gap length penalty × total length of gaps). All scores greater than or equal to 0·5 in the normalized log odds matrix were counted as matches.

RESULTS AND DISCUSSION

Sequence analysis

Fig. 1 presents a map of the insert positions for the seven cDNA clones used to analyse the BNYVV RNA 1 sequence. Clone pBF5 was obtained previously (Richards *et al.*, 1985) by the method of Heidecker & Messing (1983) with *Pst*I-linearized dT-tailed pUC9 as the first-strand primer during cDNA synthesis. The pBF5 insert has a poly(A) tail of about 100 nucleotides (nt) at one extremity, thus demonstrating that the insert extends to the 3′ terminus of the RNA molecule and defining the orientation of the sequence. Once the sequence of pBF5 was completed an oligodeoxyribonucleotide of 15 nt (primer 1, Fig. 1) complementary to the sequence near its 5′ extremity was synthesized and used to prime further cDNA synthesis, yielding clones pBF5220 and pBF51 which were in turn sequenced. The sequence was used as a guide for synthesis of a second primer (primer 2) and so on. Unfortunately, the longest cDNA inserts obtained by this procedure were only about 1200 nt in length so that it had to be carried out three times before obtaining pBF541 (Fig. 1), which extends to within 594 nt of the 5′ end (all numbering refers to the final sequence).

Initial attempts to clone cDNA covering the final 600 nt were unsuccessful. Therefore, this portion of the sequence was characterized by direct analysis of cDNA synthesized by reverse transcription from appropriate ^{32}P-labelled synthetic primers. About 300 residues of sequence could be read with confidence in this fashion so that three cycles of primer synthesis/cDNA sequence analysis, using primers 5, 6 and 7 in Fig. 1, were necessary to span the gap. The extreme 5′ terminal sequence was obtained using end-labelled cDNA extended from a primer complementary to nt 109 to 129 (primer 7 in Fig. 1). This cDNA sequence could be read with confidence up to and including the nucleotide complementary to the first residue of the RNA molecule (Bouzoubaa *et al.*, 1986). Finally, a primer identical in sequence to nt 1 to 23 of the RNA was synthesized and used to prime second-strand synthesis upon cDNA synthesized from primer 5. This cDNA was cloned (pBF553, Fig. 1) and analysis of its sequence confirmed the results provided by primer extension.

The complete sequence of BNYVV RNA-1 is 6746 residues in length (Fig. 2) which, assuming poly(A) tails of approx. 100 nt (Putz *et al.*, 1983), is within 5% of the length of 7100 nt estimated for methylmercury(II) hydroxide-denatured RNA-1 by agarose gel electrophoresis (Richards *et al.*, 1985).

Homology among the four BNYVV RNA species

Homology among the four BNYVV RNA sequences is limited to their extremities. At their 5′ extremities all four molecules begin with short runs of A, A_3 in the case of RNA-1, -2 and -4 and A_4 for RNA-3 (Fig. 3). These sequences could form part of the promoter for second-strand RNA synthesis. G residues are significantly under-represented in the first 30 or 50 residues of all four BNYVV RNA, as is also the case for the 5′ non-coding regions of several other plant viral RNA, e.g. tobacco mosaic virus (TMV) (Richards *et al.*, 1978), alfalfa mosaic virus (AMV) RNA-4 (Koper-Zwarthoff *et al.*, 1980) and carnation mottle virus (Guilley *et al.*, 1985).

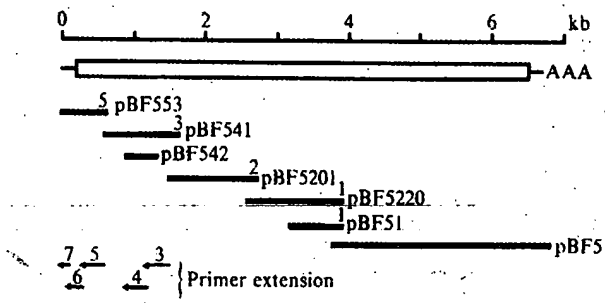


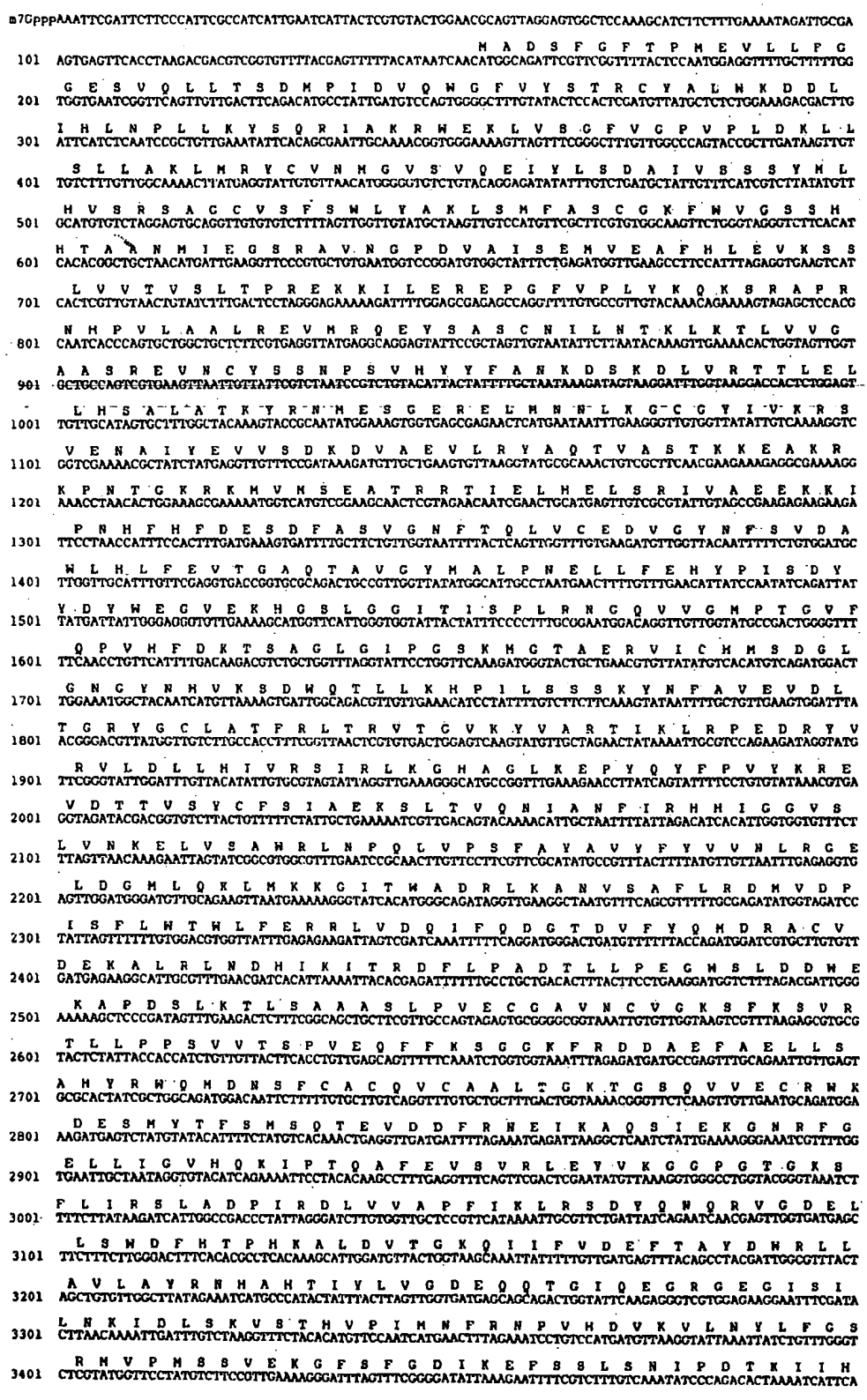
Fig. 1. Nucleotide sequencing strategy for BNYVV RNA-1 showing the alignment of the cDNA clones (solid bars) with respect to the final sequence. Regions sequenced by primer extension on RNA with synthetic oligodeoxyribonucleotide primers are indicated by arrows with the primer used in such analysis or in cDNA synthesis for cloning identified by number. The primers were complementary to the following residues of the final RNA-1 sequence: 1, nt 3845 to 3860; 2, nt 3293 to 3307; 3, nt 1527 to 1545; 4, nt 1229 to 1250; 5, nt 622 to 642; 6, nt 269 to 278; 7, nt 109 to 129. pBF542 was produced with primer 3 but the clone did not extend back to the primer sequence.

At their 3′ termini we have already shown that extensive sequence homology exists over the last 200 residues of RNA-3 and -4 (Bouzoubaa *et al.*, 1985) and over the last approx. 70 residues of RNA-2 (Bouzoubaa *et al.*, 1986). Fig. 4 depicts a sequence alignment for the 3′ terminal regions of all four RNA. The region A of RNA-2 which is homologous to much of domain A of RNA-3 and -4 is also present on RNA-1 with only three mismatches in 51 positions compared to RNA-2. Preceding this region on RNA-1 and -2 are two short homologous domains E and D and a longer region of homology F (Fig. 4). Regions A, D, E and F fall in the same order on RNA-1 and -2 but the spacing between them varies considerably. There is also homology between domain E of RNA-1/2 and a portion of domain B of RNA-3/4 as well as between portions of domains F and C (Fig. 4). Both of these alignments are of only limited extent, however, involving 13 residues (with two mismatches) for the aligned pair E/B and 12 residues with one mismatch for the pair F/C. Thus while the 3′ terminal sequences of RNAs 1 to 4 are clearly all related to one another the sequence homologies suggest that they may be further divided into two subgroups with RNA-1 and -2 forming one pair and RNA-3 and -4 the other.

For propagation, BNYVV is commonly transferred by mechanical inoculation from infected sugarbeet roots to leaves of *Chenopodium quinoa*. Isolates maintained in this host are often but not always found to lack one or both of RNA-3 and -4 or to contain deleted forms of these molecules (Richards *et al.*, 1985; Kuszala *et al.*, 1986; Burgermeister *et al.*, 1986). Thus the two smallest RNA are certainly not necessary for virus multiplication in *Chenopodium quinoa*. In infected sugarbeet roots, on the other hand, full length RNA-3 and -4 are always present (Koenig *et al.*, 1986) suggesting that RNA-3 and -4 may be essential parts of the BNYVV genome under the natural conditions of soil transmission and multiplication in roots but not under the artificial conditions of propagation in leaves. Under the latter conditions, RNA-3 and -4 could persist as satellites, undergo deletion or disappear entirely from the inoculum. In this regard, the extensive 3′ terminal sequence homology noted above strengthens the view that RNA-3 and -4 are in fact part of the BNYVV genome since such homology is like that displayed by genomic RNA for a number of multi-component viruses (Davies & Hull, 1982) but is unlike the situation for known authentic plant satellite RNA, which generally have little if any sequence homology with the helper genome (Murant & Mayo, 1982).

Coding capacity of RNA-1

RNA-1 contains a long open reading frame extending for 2109 codons from AUG(154), the first potential initiation codon in the sequence, to UAA(6481) (Fig. 2). No other open reading frame of more than 60 amino acids is present. The calculated M_r of the translation product of


m7GpppAAATTCGATTCTTCCCATTCGCCATCATTGAATCATTACTCGTGTACTGGAACGCAGTTAGGAGTGGCTCCAAAGCATCTTCTTTGAAAATAGATTGCGA
M A D S F G F T P M E V L L F G
AGTGAGTTCACCTAAGACGACGTCCGGTGTTTTACGAGTTTTTACATAATCAACATGGCAGATTCGTTCGGTTTTACTCCAATGGAGGTTTTGCTTTTTGG
G E S V Q L L T S D M P I D V Q W G F V Y S T R C Y A L W K D D L
TGGTGAATCGGTTCAGTTGTTGACTTCAGACATGCCTATTGATGTCCAGTGGGGCTTTGTATACTCCACTCGATGTTATGCTCTCTGGAAAGACGACTTG
I H L N P L L K Y S Q R I A K R W E K L V S G F V G P V P L D K L L
ATTCATCTCAATCCGCTGTTGAAATATTCACAGCGAATTGCAAAACGGTGGGAAAAGTTAGTTTCGGGCTTTGTTGGCCCAGTACCGCTTGATAAGTTGT
S L L A K L M R Y C V N M G V S V Q E I Y L S D A I V S S S Y M L
TGTCTTTGTTGGCAAAACTTATGAGGTATTGTGTTAACATGGGGGTGTCTGTACAGGAGATATATTTGTCTGATGCTATTGTTTCATCGTCTTATATGTT
H V S R S A G C V S F S W L Y A K L S M F A S C G K F W V G S S H
GCATGTGTCTAGGAGTGCAGGTTGTGTGTCTTTTAGTTGGTTGTATGCTAAGTTGTCCATGTTCGCTTCGTGTGGCAAGTTCTGGGTAGGGTCTTCACAT
H T A N N M I E G S R A V N G P D V A I S E M V E A F H L E V K S S
CACACGGCTGCTAACATGATTGAAGGTTCCCGTGCTGTGAATGGTCCGGATGTGGCTATTTCTGAGATGGTTGAAGCCTTCCATTTAGAGGTGAAGTCAT
L V V T V S L T P R E K K I L E R E P G F V P L Y K Q K S R A P R
CACTCGTTGTAACTGTATCTTTGACTCCTAGGGAGAAAAAGATTTTGGAGCGAGAGCCAGGTTTTGTGCCGTTGTACAAACAGAAAAGTAGAGCTCCACG
N M P V L A A L R E V M R Q E Y S A S C N I L N T K L K T L V V G
CAATCACCCAGTGCTGGCTGCTCTTCGTGAGGTTATGAGGCAGGAGTATTCCGCTAGTTGTAATATTCTTAATACAAAGTTGAAAACACTGGTAGTTGGT
A A S R E V N C Y S S N P S V H Y Y F A N K D S K D L V R T T L E L
GCTGCCAGTCGTGAAGTTAATTGTTATTCGTCTAATCCGTCTGTACATTACTATTTTGCTAATAAAGATAGTAAGGATTTGGTAAGGACCACTCTGGAGT
L H S A L A T K Y R N M E S G E R E L M N N L K G C G Y I V K R S
TGTTGCATAGTGCTTTGGCTACAAAGTACCGCAATATGGAAAGTGGTGAGCGAGAACTCATGAATAATTTGAAGGGTTGTGGTTATATTGTCAAAAGGTC
V E N A I Y E V V S D K D V A E V L R Y A Q T V A S T K K E A K R
GGTCGAAAACGCTATCTATGAGGTTGTTTCCGATAAAGATGTTGCTGAAGTGTTAAGGTATGCGCAAACTGTCGCTTCAACGAAGAAAGAGGCGAAAAGG
K P N T G K R K M V M S E A T R R T I E L H E L S R I V A E E K K I
AAACCTAACACTGGAAAGCGAAAAATGGTCATGTCCGAAGCAACTCGTAGAACAATCGAACTGCATGAGTTGTCGCGTATTGTAGCCGAAGAGAAGAAGA
P N H F H F D E S D F A S V G N F T Q L V C E D V G Y N F S V D A
TTCCTAACCATTTCCACTTTGATGAAAGTGATTTTGCTTCTGTTGGTAATTTTACTCAGTTGGTTTGTGAAGATGTTGGTTACAATTTTTCTGTGGATGC
W L H L F E V T G A Q T A V G Y M A L P N E L L F E H Y P I S D Y
TTGGTTGCATTTGTTCGAGGTGACCGGTGCGCAGACTGCCGTTGGTTATATGGCATTGCCTAATGAACTTTTGTTTGAACATTATCCAATATCAGATTAT
Y D Y W E G V E K H G S L G G I T I S P L R N G Q V V G M P T G V F
TATGATTATTGGGAGGGTGTTGAAAAGCATGGTTCATTGGGTGGTATTACTATTTCCCCTTTGCGGAATGGACAGGTTGTTGGTATGCCGACTGGGGTTT
Q P V H F D K T S A G L G I P G S K M G T A E R V I C H M S D G L
TTCAACCTGTTCATTTTGACAAGACGTCTGCTGGTTTAGGTATTCCTGGTTCAAAGATGGGTACTGCTGAACGTGTTATATGTCACATGTCAGATGGACT
G N G Y N H V K S D W Q T L L K H P I L S S S K Y N F A V E V D L
TGGAAATGGCTACAATCATGTTAAAAGTGATTGGCAGACGTGTGTGAAACATCCTATTTTGTCTTCTTCAAAGTATAATTTTGCTGTTGAAGTGGATTTA
T G R Y G C L A T F R L T R V T C V K Y V A R T I K L R P E D R Y V
ACGGGACGTTATGGTTGTCTTGCCACCTTTCGGTTAACTCGTGTGACTGGAGTCAAGTATGTTGCTAGAACTATAAAATTGCGTCCAGAAGATAGGTATG
R V L D L L H I V R S I R L K G H A G L K E P Y Q Y F P V Y K R E
TTCGGGTATTGGATTTGTTACATATTGTGCGTAGTATTAGGTTGAAAGGGCATGCCGGTTTGAAAGAACCTTATCAGTATTTTCCTGTGTATAAACGTGA
V D T T V S Y C F S I A E K S L T V Q N I A N F I R H H I G G V S
GGTAGATACGACGGTGTCTTACTGTTTTTCTATTGCTGAAAAATCGTTGACAGTACAAAACATTGCTAATTTTATTAGACATCACATTGGTGGTGTTTCT
L V N K E L V S A W R L N P Q L V P S F A Y A V Y F Y V V N L R G E
TTAGTTAACAAAGAATTAGTATCGGCGTGGCGTTTGAATCCGCAACTTGTTCCTTCGTTCGCATATGCCGTTTACTTTTATGTTGTTAATTTGAGAGGTG
L D G M L Q K L M K K G I T W A D R L K A N V S A F L R D M V D P
AGTTGGATGGGATGTTGCAGAAGTTAATGAAAAAGGTATCACATGGGCAGATAGGTTGAAGGCTAATGTTTCAGCGTTTTTGCGAGATATGGTAGATCC
I S F L N T W L F E R R L V D Q I F Q D G T D V F Y Q M D R A C V
TATTAGTTTTTTGTGGACGTGGTTATTTGAGAGAAGATTAGTCGATCAAATTTTTCAGGATGGGACTGATGTTTTTTACCAGATGGATCGTGCTTGTGTT
D E K A L R L N D H I K I T R D F L P A D T L L P E G W S L D D W E
GATGAGAAGGCATTGCGTTTGAACGATCACATTAAAATTACACGAGATTTTTTGCCTGCTGACACTTTACTTCCTGAAGGATGGTCTTTAGACGATTGGG
K A P D S L K T L S A A A S L P V E C C G A V N C V G K S F K S V R
AAAAAGCTCCCGATAGTTTGAAGACTCTTTCGGCAGCTGCTTCGTTGCCAGTAGAGTGCGGGGCGGTAAATTGTGTTGGTAAGTCGTTTAAGAGCGTGCG
T L L P P S V V T S P V E Q F F K S G G K F R D D A E F A E L L S
TACTCTATTACCACCATCTGTTGTTACTTCACCTGTTGAGCAGTTTTTCAAATCTGGTGGTAAATTTAGAGATGATGCCGAGTTTGCAGAATTGTTGAGT
A H Y R W Q M D N S F C A C Q V C A A L T G K T G S Q V V E C R N K
GCGCACTATCGCTGGCAGATGGACAATTCTTTTTGTGCTTGTCAGGTTTGTGCTGCTTTGACTGGTAAAACCGGGTTCTCAAGTTGTTGAATGCAGATGGA
D E S M Y T F S M S Q T E V D D F R N E I K A Q S I E K G N R F G
AAGATGAGTCTATGTATACATTTTCTATGTCACAAACTGAGGTTGATGATTTTAGAAATGAGATTAAGGCTCAATCTATTGAAAAGGGAAATCGTTTTGG
E L L I G V H Q K I P T Q A F E V S V R L E Y V K G G P G T G K S
TGAATTGCTAATAGGTGTACATCAGAAAATTCCTACACAAGCCTTTGAGGTTTCAGTTCGACTCGAATATGTTAAAGGTGGGCCTGGTACGGGTAAATCT
F L I R S L A D P I R D L V V A P F I K L R S D Y Q W Q R V G D E L
TTTCTTATAAGATCATTGGCCGACCCTATTAGGGATCTTGTGGTTGCTCCGTTCATAAAATTGCGTTCTGATTATCAGAATCAACGAGTTGGTGATGAGC
L S W D F H T P H K A L D V T G K Q I I F V D E F T A Y D W R L L
TTCTTTCTTGGGACTTTCACACGCCTCACAAAGCATTGGATGTTACTGGTAAGCAAATTATTTTTGTTGATGAGTTTACAGCCTACGATTGGCGTTTACT
A V L A Y R N H A H T I Y L V G D E Q Q T G I Q E G R G E G I S I
AGCTGTGTTGGCTTATAGAAATCATGCCCATACTATTTACTTAGTTGGTGATGAGCAGCAGACTGGTATTCAAGAGGGTCGTGCAGAAGGAATTTCGATA
L N K I D L S K V S T H V P I M N F R N P V H D V K V L N Y L F G S
CTTAACAAAATTGATTTGTCTAAGGTTTCTACACATGTTCCAATCATGAACTTTAGAAATCCTGTCCATGATGTTAAGGTATTAAATTATCTGTTTGGGT
R N V P M S S V E K G F S F G D I K E F S S L S N I P D T K I I H
CTCGTATGGTTCCTATGTCTTCCGTTGAAAAGGGATTTAGTTTCGGGGATATTAAAGAATTTTCGTCTTTGTCAAATATCCCAGACACTAAAATCATTCA

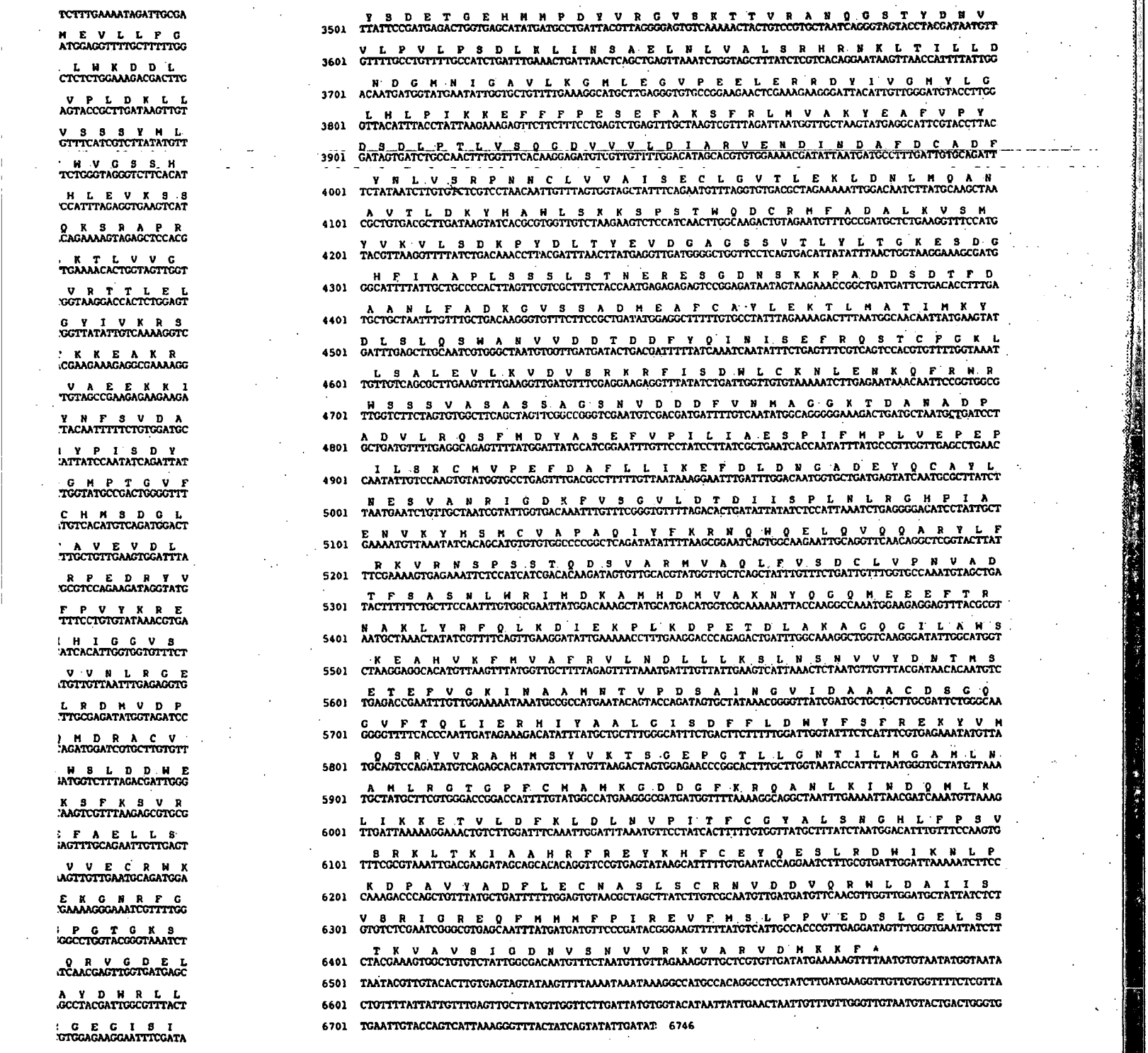


Fig. 2. Nucleotide sequence of BNYVV RNA-1. The sequence is written as DNA and the amino acid sequence of the 237K open reading frame (see text) is indicated in one-letter code above the sequence with an asterisk denoting a termination codon. The 3' terminal poly(A) tail is not shown.

RNA

1 cap-AAAUUCGAUUCUUCCCAUUCGCCAUCAUUG

2 cap-AAAUUCUAACUAUUAUCUCCAUUGAAUAGA

3 cap-AAAAUUCAAAAUUUACCAUUACAUAUUGGU

4 cap-AAAUCAAAUCUCAAAUAUAUAUUUGUAUUU

Fig. 3. Sequence at the 5′ terminus of the four BNYVV RNA. A filled circle has been placed above each G residue.

this long open reading frame, assuming translation begins with the first AUG, is 237 389 (237K). The coding capacity of all four BNYVV RNA is depicted in Fig. 5.

Computer-assisted sequence comparisons with proteins encoded by several RNA viruses reveal domains of homology common to non-structural proteins thought to be involved in viral replication (Haseloff *et al.*, 1984; Kamer & Argos, 1984; Ahlquist *et al.*, 1985; Ali Rezaian *et al.*, 1984, 1985). The BNYVV 237K polypeptide also contains sequences of this sort. Fig. 6(*a* and *b*) shows portions of the TMV P126 protein and the overlapping readthrough protein P183 aligned with corresponding parts of the BNYVV 237K polypeptide. The location of the homologous domains on the genetic map of each virus is given in Fig. 6(*c*). Note that the regions shown (T1 and T2 of TMV, B1 and B2 of BNYVV) are those of greatest homology; the alignments can be extended in both directions from these core sequences although with less confidence. Haseloff *et al.* (1984) and Ahlquist *et al.* (1985) have aligned the TMV T1 and T2 regions with the RNA-1- and -2-encoded proteins of the tripartite viruses AMV and brome mosaic virus (BMV) and with portions of the Sindbis virus p270 non-structural polyprotein. Kamer & Argos (1984) have extended the set of homologies characteristic of the readthrough portion of the TMV P183 protein (T2) to include cowpea mosaic virus and the picornaviruses. These comparisons provide a consensus alignment in which certain residues are conserved in virtually all members of the set. Fig. 6 shows that these consensus residues are with few exceptions also present in the aligned BNYVV sequence, thus suggesting involvement of the BNYVV 237K polypeptide in viral RNA replication and a certain generality as to the mechanism. It should be noted, however, that homology also exists between the N-terminal regions of the Sindbis virus P270 polyprotein, TMV P126 and the AMV and BMV RNA-2 proteins (Ahlquist *et al.*, 1985). A counterpart of this domain in the BNYVV 237K polypeptide, if it exists, is below the level of homology readily detectable by the search methods used.

Nothing is yet known concerning the translation strategy of BNYVV RNA-1 *in vivo* but in a previous cell-free translation study using rabbit reticulocyte lysate, RNA-1 was found to direct synthesis of 50K and 150K polypeptides as well as lesser amounts of an approx. 200K species (Ziegler *et al.*, 1985). In view of the sequence results it appears probable that these three polypeptides all have the same N terminus with the shorter products arising by premature termination or ribosome 'stalling' (Lindhout *et al.*, 1985) although alternative interpretations such as internal initiation cannot be conclusively ruled out. When translated in a wheat germ system, RNA-1 directs synthesis of comparable amounts of two very long polypeptides, one of about 200K and one which is slightly longer (data not shown). The latter species could correspond to translation of the entire RNA-1 open reading frame.

Homology between BNYVV RNA-2 and barley stripe mosaic virus RNA-β

Barley stripe mosaic virus (BSMV) is a tripartite virus for which, like BNYVV, synthesis of coat protein is directed by the second largest genomic RNA (RNA-β) (Gustafson & Armour, 1986). The complete nucleotide sequence of RNA-β from the type strain of BSMV has recently been reported (Gustafson & Armour, 1986). The sequence contains four open reading frames

laced above each

is 237389 (237K).

ral RNA viruses
involved in viral
Ali Rezaian *et al.*,
rt. Fig. 6(*a* and *b*)
tein P183 aligned
f the homologous
egions shown (T1
lignments can be
dence. Haseloff *et*
with the RNA-1-
(BMV) and with
rgos (1984) have
f the TMV P183
parisons provide
l members of the
ent in the aligned
ide in viral RNA
d, however, that
270 polyprotein,
A counterpart of
homology readily

-1 *in vivo* but in a
is found to direct
ox. 200K species
that these three
ig by premature
e interpretations
in a wheat germ
/peptides, one of
er species could

NA-β
VV, synthesis of
fson & Armour,
MV has recently
reading frames

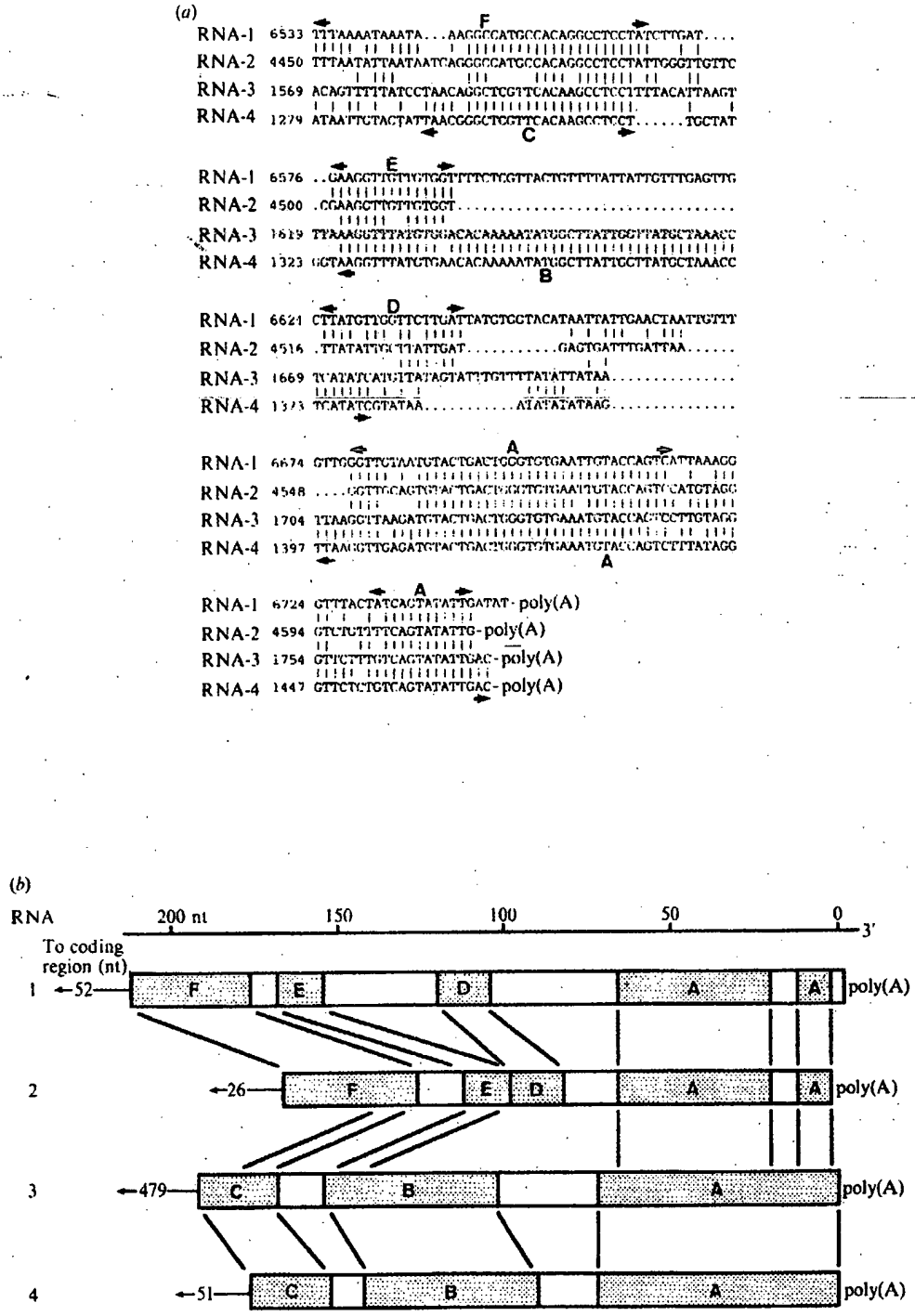


Fig. 4. A BESTFIT alignment of the 3′ terminal sequence of the four BNYVV RNA (*a*) and a schematic representation of the 3′ terminal region (*b*). The homologous blocks of sequence denoted by letters in (*a*) are stippled and identified in (*b*).

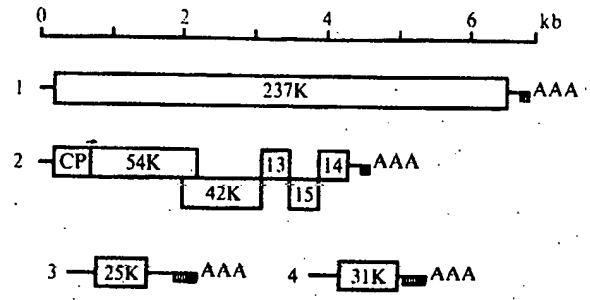


Fig. 5. Genetic maps of the four BNYVV RNA. Open reading frames are depicted as hollow rectangles with the coding capacity of each indicated within. CP refers to the 21K viral coat protein. The small arrow above RNA-2 represents a leaky termination codon (Ziegler *et al.*, 1985). Solid bars represent 3' terminal homologous sequences common to all four RNA and the hatched bars the position of 3' terminal homologous sequences found only in RNA-3 and -4.

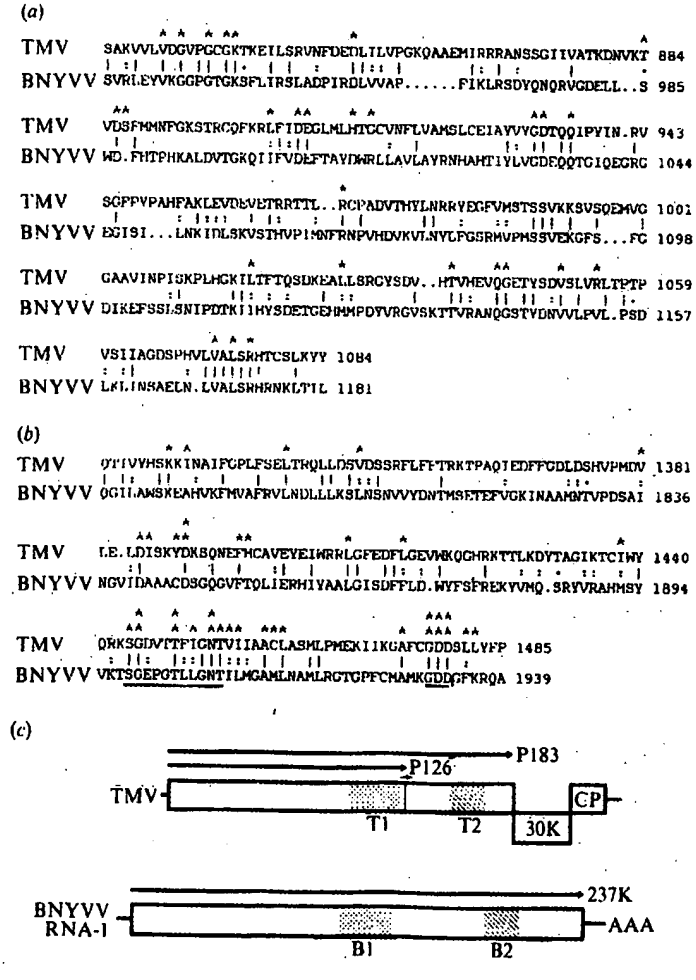


Fig. 6

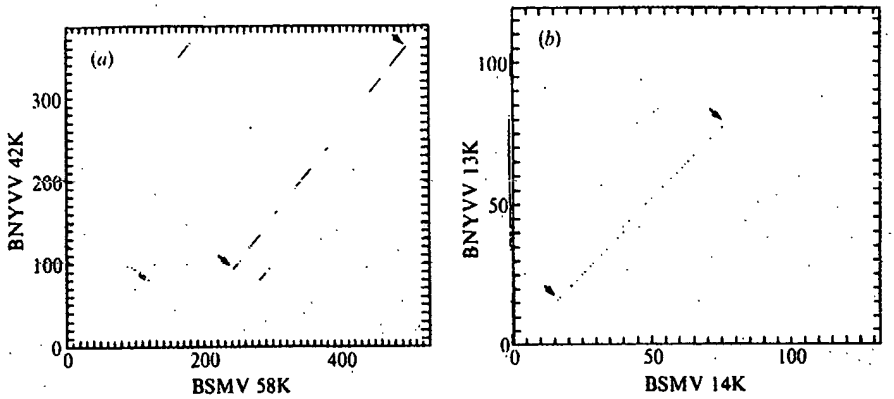


Fig. 7. Graphic homology comparisons of (*a*) BSMV 58K and BNYVV 42K proteins and (*b*) BSMV 14K and BNYVV 13K proteins. Window, 30; stringency, 32·7 in (*a*) and 32·5 in (*b*). Arrows mark the limits of the homologous regions.

s hollow rectangles
protein. The small
id bars represent 3′
the position of 3′

884
985
943
1044
1001
1098
1059
1157
1381
1836
1440
1894

which code for viral coat protein, and polypeptides of 58K, 14K and 17K. The coat protein cistron lies nearest the 5′ terminus of the RNA, the cistron for the 58K polypeptide occupies the central portions of the RNA and the 17K coding region is nearest the 3′ end (Fig. 8*c*). The cistron for the 14K polypeptide overlaps the 3′ end of the 58K cistron, the 5′ portion of the 17K cistron and the region which separates them (Fig. 8*c*).

In view of the superficial resemblance in genetic organization between BNYVV RNA-2 and BSMV RNA-β, homology dot plots were generated for all possible pairwise combinations of the proteins encoded by the two RNA. Such analysis reveals extensive amino acid homology between the BNYVV 42K polypeptide and the C-terminal half of the BSMV 58K polypeptide, and between the BNYVV 13K and BSMV 14K proteins (Fig. 7). Detailed alignments are shown in Fig. 8(*a*) and (*b*). These alignments are statistically highly significant according to the empirical test of Doolittle (1981). The quality scores for the 42K/58K (Fig. 8*a*) and 13K/14K (Fig. 8*b*) alignments are, respectively, 17·7 and 24·5 standard deviations above the means of the quality scores obtained by reiterative comparisons of the same sequences after randomization. A difference of more than 3 standard deviations in this test is normally accepted as indicating statistical significance (Doolittle, 1981).

No information is yet available for either virus concerning the roles these homologous polypeptides might play in the virus multiplication cycle but the very existence of such homologies provides evidence that the open reading frames in question do in fact correspond to expressed genes. Such assurance is particularly comforting for the short 13K BNYVV and 14K BSMV coding regions. The homologies would also seem to indicate that BNYVV and BSMV are more closely related than hitherto suspected.

Fig. 6. BESTFIT alignments of regions T1 and B1 (*a*) and T2 and B2 (*b*) of the TMV and BNYVV RNA-1 large non-structural proteins. Identical aligned amino acids are indicated by vertical lines and favoured substitution [score 0·5 in the normalized log odds matrix (see Methods)] by two dots. Single dots indicate alignments of S and T. Closely similar alignments were generated using a direct identity scoring matrix. In (*a*) asterisks above the sequence refer to residues conserved in the alignment of the TMV sequence with that of TMV, AMV, BMV and Sindbis virus (Ahlquist *et al.*, 1985). In (*b*) double asterisks denote strictly conserved residues and single asterisks conserved familial residues in the alignment of Kamer & Argos (1984). The underlined sequences indicate sequence motifs conserved in the sequence of the putative replicases of all plus-strand RNA viruses which have been sequenced to date (Guilley *et al.*, 1985 and further observations). (*c*) Genetic organization of TMV RNA and BNYVV RNA-1. The positions of the homologous domains T1 and T2 on TMV and B1 and B2 on BNYVV RNA-1 are indicated. The small arrow represents the leaky termination codon whose readthrough gives rise to P183 (Pelham, 1978).

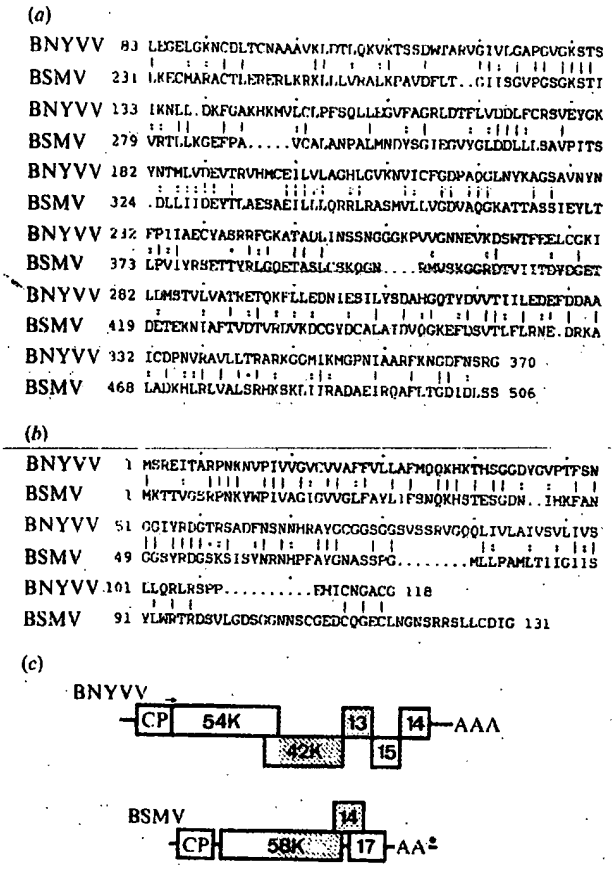


Fig. 8. BESTFIT alignments of portions of the BNYVV 42K protein with the BSMV 58K protein (*a*) and of the BNYVV 13K protein with the BSMV 14K protein (*b*). Symbols for amino acid matches are as in Fig. 6(*c*). (*c*) Genetic maps of BNYVV RNA-2 and BSMV RNA-β. The homologous regions shown in (*a*) are shaded. The small arrow above the BNYVV map represents a leaky termination codon (Bouzoubaa *et al.*, 1986). The star in the BSMV map indicates the position of the tRNA-like structure following the internal poly(A) tract (Gustafson & Armour, 1986).

CONCLUSIONS

With this report, BNYVV becomes the eighth plant RNA virus for which complete sequence information is available. In genome organization BNYVV shares features with certain other plant viruses while other properties set it apart. In particular, BNYVV joins the tobraviruses (Harrison & Robinson, 1978), soil-borne wheat mosaic virus (SBWMV) (Hsu & Brakke, 1985) and Indian peanut clump virus (IPCV) (Mayo & Reddy, 1985) in expressing its coat protein as a primary translation product of its second largest RNA. In both BNYVV (Ziegler *et al.*, 1985) and SBWMV (Hsu & Brakke, 1985) the coat protein cistron can be read through to produce longer polypeptides. BNYVV RNA are 3′ polyadenylated, however, but those of SBWMV, IPCV, BSMV and the tobraviruses are not (Harrison & Robinson, 1978; Kozlov *et al.*, 1984; Hsu & Brakke, 1985; Mayo & Reddy, 1985). Furthermore, SBWMV, IPCV and the tobraviruses have bipartite genomes (Harrison & Robinson, 1978; Shirako & Brakke, 1984; Reddy *et al.* 1985) but, as noted above, the genome of BNYVV may well consist of four components, at least when propagated in its natural host.

We have shown in the preceding section that there are similarities in overall genetic organization between BNYVV RNA-2 and RNA-β of BSMV as well as homologies at the amino acid sequence level. The other RNA components of BSMV, however, differ significantly from their counterparts in BNYVV: (i) BSMV RNA-α, although it resembles BNYVV RNA-1 in directing synthesis of a long polypeptide (Dolja *et al.*, 1983), is only about half the length of BNYVV RNA-1; (ii) BSMV RNA-γ is dicistronic, encoding 75K to 85K and 17K polypeptides, the latter expressed from a subgenomic RNA (Gustafson *et al.*, 1981; Dolja *et al.*, 1983; Jackson *et al.*, 1983). Thus BNYVV appears to possess a unique spectrum of characteristics distinguishing it from all known groups of plant RNA viruses.

The technical assistance of Miss D. Beck is gratefully acknowledged while C. Putz and his colleagues (INRA, Colmar) were a source of advice, encouragement and virus. S.B. benefited from a grant from the Institut Technique Français de la Betterave Industrielle.

REFERENCES

AHLQUIST, P., STRAUSS, E. G., RICE, C. M., STRAUSS, J. M., HASELOFF, J. & ZIMMERN, D. (1985). Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. *Journal of Virology* **53**, 536–542.

ALI REZAIAN, M., WILLIAMS, R. H. V., GORDON, K., GOULD, A. & SYMONS, R. H. (1984). Nucleotide sequence of cucumber-mosaic-virus RNA 2 reveals a translation product significantly homologous to corresponding proteins of other viruses. *European Journal of Biochemistry* **143**, 277–284.

ALI REZAIAN, M., WILLIAMS, R. H. V. & SYMONS, R. H. (1985). Nucleotide sequence of cucumber-mosiac-virus RNA-1; presence of a sequence complementary to part of the viral satellite RNA and homologies with other viral RNAs. *European Journal of Biochemistry* **150**, 331–339.

BOUZOUBAA, S., GUILLEY, H., JONARD, G., RICHARDS, K. & PUTZ, C. (1985). Nucleotide sequence analysis of RNA-3 and RNA-4 of beet necrotic yellow vein virus, isolates F2 and G1. *Journal of General Virology* **66**, 1553–1564.

BOUZOUBAA, S., ZIEGLER, V., BECK, D., GUILLEY, H., RICHARDS, K. & JONARD, G. (1986). Nucleotide sequence of beet necrotic yellow vein virus RNA-2. *Journal of General Virology* **67**, 1689–1700.

BURGERMEISTER, W., KOENIG, R., WEICH, H., SEBALD, W. & LESEMANN, D. E. (1986). Diversity of the RNAs in thirteen isolates of beet necrotic yellow vein virus in *Chenopodium quinoa* detected by means of cloned cDNAs. *Journal of Phytopathology* **115**, 229–242.

DAVIES, J. W. & HULL, R. (1982). Genome expression of plant positive-strand RNA viruses. *Journal of General Virology* **61**, 1–14.

DAYHOFF, M. O., SCHWARTZ, R. M. & ORCUTT, B. C. (1978). A model for evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, vol. 5, supplement 3, pp. 345–352. Edited by M. O. Dayhoff. Washington, D.C.: NBR Foundation, Georgetown University.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

DOLJA, V. V., LUNIA, N. A., LEISER, R. M., STANARIUS, T., BELZHELARSKAYER, S. N., KOVLOV, Y. V. & ATABEKOV, J. G. (1983). A comparative study of the in vitro translation products of individual RNAs from two-, three- and four-component strains of barley stripe mosaic virus. *Virology* **127**, 1–14.

DOOLITTLE, R. F. (1981). Similar amino acid sequences: chance or common ancestry? *Science* **214**, 149–159.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

GUILLEY, H., CARRINGTON, J. C., BALÁZS, E., JONARD, G., RICHARDS, K. & MORRIS, T. J. (1985). Nucleotide sequence and genome organization of carnation mottle virus RNA. *Nucleic Acids Research* **13**, 6663–6677.

GUSTAFSON, G. & ARMOUR, S. L. (1986). The complete nucleotide sequence of RNAβ from the type strain of barley stripe mosaic virus. *Nucleic Acids Research* **14**, 3895–3909.

GUSTAFSON, G. D., LARKINS, B. A. & JACKSON, A. O. (1981). Comparative analysis of polypeptides synthesized *in vivo* and *in vitro* by two strains of barley stripe mosaic virus. *Virology* **111**, 579–587.

HARRISON, B. D. & ROBINSON, D. J. (1978). The Tobraviruses. *Advances in Virus Research* **23**, 25–77.

HASELOFF, J., GOELET, P., ZIMMERN, D., AHLQUIST, P., DASGUPTA, R. & KAESBERG, P. (1984). Striking similarities among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. *Proceedings of the National Academy of Sciences, U.S.A.* **81**, 4358–4362.

HEIDECKER, G. & MESSING, J. (1983). Sequence analysis of zein cDNAs obtained by an efficient mRNA cloning method. *Nucleic Acids Research* **11**, 4891–4906.

HSU, Y. H. & BRAKKE, M. K. (1985). Cell-free translation of soil-borne wheat mosaic virus RNAs. *Virology* **143**, 272–279.

JACKSON, A. O., DAWSON, J. R. O., COVEY, S. N., HULL, R., DAVIES, J. W., MACFARLAND, J. E. & GUSTAFSON, G. D. (1983). Sequence relations and coding properties of a subgenomic RNA isolated from barley stripe mosaic virus. *Virology* **127**, 37–44.

KAMER, G. & ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269–7283.

KOENIG, R., BURGERMEISTER, W., WEICH, H., SEBALD, W. & KOTHE, C. (1986). Uniform RNA patterns of beet necrotic yellow vein virus in sugarbeet roots, but not in leaves from several plant species. *Journal of General Virology* **67**, 2043–2046.

KOPER-ZWARTHOFF, E., BREDERODE, F., VEENEMAN, G., VAN BOOM, J. & BOL, J. (1980). Nucleotide sequences at the 5′ termini of the alfalfa mosaic virus RNAs and the intercistronic junction in RNA 3. *Nucleic Acids Research* **8**, 5635–5647.

KOZLOV, Y. V., RUPASOV, V. V., ADYSHEV, D. M., BELGELARSKAYA, S. N., AGRANOVSKY, A. A., MANKIN, A. S., MOROZOV, S. Y. & ATABEKOV, J. G. (1984). Nucleotide sequence at the 3′-terminal tRNA-like structure in barley stripe mosaic virus. *Nucleic Acids Research* **12**, 4001–4009.

KUSZALA, M., ZIEGLER, V., BOUZOUBAA, S., RICHARDS, K., PUTZ, C., GUILLEY, H. & JONARD, G. (1986). Beet necrotic yellow vein virus: different isolates are serologically similar but differ in RNA composition. *Annals of Applied Biology* (in press).

LINDHOUT, P., NEELEMAN, L., VAN TOL, H. & VAN VLOTEN-DOTING, L. (1985). Ribosomes are stalled during *in vitro* translation of alfalfa mosaic virus RNA 1. *European Journal of Biochemistry* **152**, 625–631.

MAIZEL, J. & LENK, R. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proceedings of the National Academy of Sciences, U.S.A.* **78**, 7665–7669.

MAXAM, A. M. & GILBERT, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 560–564.

MAYO, M. A. & REDDY, D. V. R. (1985). Translation products of RNA from Indian peanut clump virus. *Journal of General Virology* **66**, 1347–1351.

MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20–77.

MURANT, A. F. & MAYO, M. A. (1982). Satellites of plant viruses. *Annual Review of Phytopathology* **20**, 49–70.

OKAYAMA, Y. & BERG, P. (1982). High-efficiency cloning of full-length cDNA. *Molecular and Cellular Biology* **2**, 161–170.

PELHAM, H. R. B. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature, London* **272**, 469–471.

PUTZ, C. (1977). Composition and structure of beet necrotic yellow vein virus. *Journal of General Virology* **35**, 397–401.

PUTZ, C., PINCK, L., PINCK, M. & FRITSCH, C. (1983). Identification of 3′ and 5′ ends of beet necrotic yellow vein virus RNAs. Presence of poly A sequences. *FEBS Letters* **156**, 41–56.

REDDY, D. V. R., ROBINSON, D. J., ROBERTS, I. M. & HARRISON, B. D. (1985). Genome properties and relationships of Indian peanut clump virus. *Journal of General Virology* **66**, 2011–2016.

RICHARDS, K. E., GUILLEY, H., JONARD, G. & HIRTH, L. (1978). Nucleotide sequence of the 5′ extremity of tobacco mosaic virus RNA. *European Journal of Biochemistry* **84**, 513–519.

RICHARDS, K., JONARD, G., GUILLEY, H., ZIEGLER, V. & PUTZ, C. (1985). *In vitro* translation of beet necrotic yellow vein virus RNA and studies of sequence homology among the RNA species using cloned cDNA probes. *Journal of General Virology* **66**, 345–350.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

SHIRAKO, Y. & BRAKKE, M. K. (1984). Two purified RNAs of soil-borne wheat mosaic virus are needed for infection. *Journal of General Virology* **65**, 119–127.

STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951–2961.

TAMADA, T. (1975). Beet necrotic yellow vein virus. *Commonwealth Mycological Institute/Association of Applied Biologists Descriptions of Plant Viruses*, no. 144.

ZIEGLER, V., RICHARDS, K., GUILLEY, H., JONARD, G. & PUTZ, C. (1985). Cell-free translation of beet necrotic yellow vein virus: readthrough of the coat protein cistron. *Journal of General Virology* **66**, 2079–2087.

(Received 20 August 1986)